*INTERSPEECH 2017*
August 20–24, 2017, Stockholm, Sweden

# Wavelet Speech Enhancement Based on Robust Principal Component Analysis

*Chia-Lung Wu*[1], *Hsiang-Ping Hsu*[1], *Syu-Siang Wang*[2], *Jeih-Weih Hung*[3], *Ying-Hui Lai*[4], *Hsin-Min Wang*[5], *and Yu Tsao*[2]

[1]Investigation Bureau, Ministry of Justice, R.O.C
[2]Research Center for Information Technology Innovation, Academia Sinica, R.O.C
[3]Dept of Electrical Engineering, National Chi Nan University, R.O.C
[4]Department of Electrical Engineering, Yuan Ze University, R.O.C
[5]Institute of Information Science, Academia Sinica, R.O.C

sypdbhee@citi.sinica.edu.tw, yu.tsao@citi.sinica.edu.tw

## Abstract

Most state-of-the-art speech enhancement (SE) techniques prefer to enhance utterances in the frequency domain rather than in the time domain. However, the overlap-add (OLA) operation in the short-time Fourier transform (STFT) for speech signal processing possibly distorts the signal and limits the performance of the SE techniques. In this study, a novel SE method that integrates the discrete wavelet packet transform (DWPT) and a novel subspace-based method, robust principal component analysis (RPCA), is proposed to enhance noise-corrupted signals directly in the time domain. We evaluate the proposed SE method on the Mandarin hearing in noise test (MHINT) sentences. The experimental results show that the new method reduces the signal distortions dramatically, thereby improving speech quality and intelligibility significantly. In addition, the newly proposed method outperforms the STFT-RPCA-based speech enhancement system.

**Index Terms**: short-time Fourier transform, discrete wavelet packet transform, robust principal component analysis, speech enhancement

## 1. Introduction

The goal of speech enhancement (SE) techniques is to extract clean speech signal from noisy input to improve sound quality and intelligibility simultaneously, thereby improving speech interactions for various acoustic applications in the presence of noise [1, 2]. In general, conventional spectral-based SE techniques can be divided into spectral subtractive, statistical-model, and subspace-based categories [3], with a simple assumption that noisy signals are constructed by adding noise to clean utterances. The algorithms for spectral subtraction basically subtract the estimated noise components from the noise-corrupted speech spectra. Some well-known spectral subtractive schemes include the Wiener filter [4] and geometric approach to spectral subtraction (GSS) [5]. The school of statistical-model SE methods aims to minimize a specific distortion measure between the original noise-free signal and the enhanced counterpart. Representative methods for this include minimum-mean-square-error short-time spectral amplitude estimation (MMSE-STSA) [6] and the generalized maximum a posteriori spectral amplitude (GMAPA) estimation [7]. The primary idea of subspace-based approaches is to split a noisy signal into two subspaces, one for clean signal and the other for noise, and then minimizes the noise component in the clean-signal subspace. Some well-developed subspace techniques include singular value decomposition (SVD) [8], a generalized subspace approach with built-in prewhitening [9], and principal component analysis (PCA) [10–13].

Most of the state-of-the-art SE techniques enhance the short-time Fourier transform (STFT)-based spectral magnitude, while leaving the input noise-corrupted phase components unchanged, which may distort the enhanced speech signal and limit the SE performance [14]. Accordingly, discrete wavelet transform (DWT)-based SE approaches have been developed [14–17] to avoid the aforementioned distortion issue by processing time signals directly, and achieving a better SE performance. To extend the previous spectral subtractive work of [14], this paper proposes a novel subspace SE scheme that improves the robust principal component analysis (RPCA)-based SE method [18] using discrete wavelet packet transform (DWPT) [19,20]. The resulting new SE system, termed DWPT-RPCA SE for simplicity, can enhance the signal in the time domain directly.

In the proposed DWPT-RPCA SE system, DWPT is first applied to decompose a full-band time signal into sub-bands. For each sub-band signal, a signal space is constructed by framing the sequence. Next, RPCA is applied to create a gain subspace from the original signal space. The overlap-add (OLA) operation is then used for this virtually gain subspace to recover the gain function to enhance the corresponded sub-band signal. Finally, the inverse DWPT (IDWPT) is performed to all enhanced sub-band sequences to reconstruct the enhanced full-band time signal. The performance of the DWPT-RPCA SE was evaluated on the Mandarin hearing in noise test (MHINT) sentences. The experimental results confirm that the proposed DWPT-RPCA SE outperforms the STFT-RPCA SE and significantly improves speech quality and intelligibility in noise-corrupted situations.

The rest of the paper is organized as follows. Techniques of DWPT and RPCA are briefly introduced in Sections 2 and 3, respectively. Section 4 presents the proposed DWPT-RPCA SE framework. Experiments and the respective analysis are given in Section 5 to demonstrate the performance of the proposed approach. Section 6 summarizes our findings.

## 2. Discrete Wavelet Packet Transform

A set of well-defined low- and high-pass filters together with an factor-2 down/up-sampling process are performed for DWPT/IDWPT serving as a distortion-less analysis/synthesis for an arbitrary signal. An example of the DWPT/IDWPT with a two-level analysis/synthesis ($J = 2$) is given on the left/right side of Fig. 1. For the left side, a full-band time signal $\mathbf{n}_0^0$ is first decomposed into $\mathbf{n}_1^1$ and $\mathbf{n}_2^1$ sub-band signals carrying informa-



http://dx.doi.org/10.21437/Interspeech.2017-781

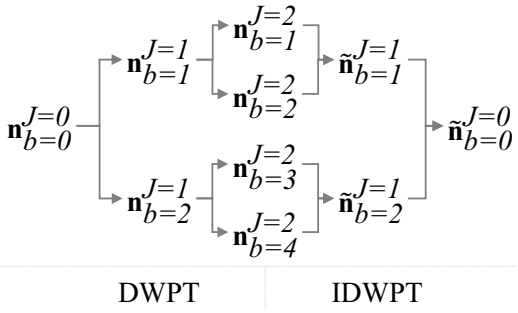



Figure 1: *An example of level-2 ($J = 2$) DWPT and IDWPT.*

tion for the low- and high-frequency components, respectively. The decomposition operation is applied again to each of the two sub-band signals, and four sub-band signals are generated. The DWPT is formulated in Eq. (1):

$$\begin{aligned} &\mathbf{n}_b^J = DWPT_b^J\{\mathbf{n}_0^0\}, \quad b = 1, 2, 3 \cdots, 2^J, \\ &\mathbf{n}^J = \{\mathbf{n}_1^J, \mathbf{n}_2^J, \cdots, \mathbf{n}_{2^J}^J\}, \end{aligned} \tag{1}$$

where $\mathbf{n}^J$ represents the whole set of $\mathbf{n}_b^J$, which is denoted as any sub-band signal produced by DWPT. In addition, $J$ is the level of DWPT, and $b$ is the sub-band index.

On the right side of Fig. 1, IDWPT integrates four decomposed sub-band signals, resulting in two sub-band signals, $\tilde{\mathbf{n}}_1^1$ and $\tilde{\mathbf{n}}_2^1$. The same IDWPT process is applied to generate a full-band time signal. The IDWPT is formulated in Eq. (2):

$$\tilde{\mathbf{n}}_0^0 = IDWPT^J\{\mathbf{n}^J\}. \tag{2}$$

Furthermore, Daubechies order 10 wavelet bases ('db10') were performed to design the DWPT filter set in the paper.

Please note that the reconstructed signal is identical to the input one ($\tilde{\mathbf{n}}_0^0 = \mathbf{n}_0^0$) through a perfect transformation of DWPT/IDWPT with a well-defined filter sets. In addition, all the filter-based decompositions and reconstructions are performed in time domain. For more details, please refer to [19] and [20].

## 3. Robust Principal Component Analysis

Consider a noisy input matrix $\mathbf{N} = \mathbf{C} + \mathbf{E}$, in which $\mathbf{C}$ and $\mathbf{E} \in \mathcal{R}^{L\times K}$ denote the $L$-dimensional clean-signal and noise matrices, respectively. Assuming the clean speech is a low-rank structure, RPCA can be adopted to decompose the noisy input $\mathbf{N}$ into clean-signal and noise subspaces. The constrained optimization problem for RPCA is formulated in Eq. (3) [21]:

$$\min\{\|\mathbf{E}\|_* + \lambda\|\mathbf{C}\|_1\}, \qquad \text{subject to} \quad \mathbf{N} - \mathbf{C} - \mathbf{E} = \mathbf{0}, \tag{3}$$

where $\|\mathbf{E}\|_*$ represents the nuclear norm and is determined as the sum of the singular values of $\mathbf{E}$, and $\lambda$ is a scalar factor. Inexact augmented Lagrange multipliers [21] are applied for solving Eq. (3) with the Lagrangian function as in Eq. (4):

$$\begin{aligned} &L(\mathbf{C}, \mathbf{E}, \mathbf{Y}, \mu) = \\ &\|\mathbf{E}\|_* + \lambda\|\mathbf{C}\|_1 + \langle \mathbf{Y}, \mathbf{N} - \mathbf{C} - \mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{N} - \mathbf{C} - \mathbf{E}\|_F^2, \end{aligned} \tag{4}$$

where $\|\cdot\|_F$ and $\langle\cdot\rangle$ are the Frobenius norm and inner product operators, respectively, $\mathbf{Y}$ is an auxiliary matrix and $\mu$ is a scalar multiplier. In addition, the initial $\mathbf{Y}$ is set to the noisy input matrix $\mathbf{N}$. The minimization of the function $L$ in Eq. (4) is obtained by implementing the following process iteratively:

$$\begin{cases} \mathbf{E}_{k+1} \leftarrow \arg\min_{\mathbf{E}} L(\mathbf{C}_k, \mathbf{E}, \mathbf{Y}_k, \mu_k), \\ \mathbf{C}_{k+1} \leftarrow \arg\min_{\mathbf{C}} L(\mathbf{C}, \mathbf{E}_{k+1}, \mathbf{Y}_k, \mu_k), \\ \mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k(\mathbf{N} - \mathbf{C}_{k+1} - \mathbf{E}_{k+1}), \\ k \leftarrow k + 1, \end{cases} \tag{5}$$

where $k$ is the iteration index.

Notably, the RPCA-derived clean-signal and noise matrices are further used to form a gain matrix $\mathbf{G}$ as in Eq. (6),

$$\mathbf{G} = \mathbf{C}./(\mathbf{C} + \mathbf{E}), \tag{6}$$

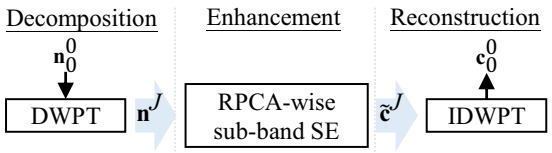
which helps to create the enhanced signal matrix $\tilde{\mathbf{C}}$ by

$$\tilde{\mathbf{C}} = \mathbf{N}. \times \mathbf{G}, \tag{7}$$

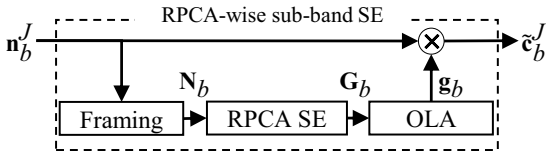
where "./" in Eq. (6) and ".×" in Eq. (7) denote the element-wise division and multiplication operators, respectively.

## 4. DWPT-RPCA Speech Enhancement System

The concept of the newly proposed DWPT-RPCA SE is to apply RPCA-wise enhancement to the DWPT-based sub-band time signal. The flowchart for DWPT-RPCA is depicted in Fig. 2. Fig. 2 (a) shows the block diagram of DWPT-RPCA, while the detailed sub-band-based RPCA SE is illustrated in Fig. 2 (b).




(a) Block diagram of DWPT-RPCA




(b) Flowchart of RPCA-based subband SE

Figure 2: *The flowchart of the DWPT-RPCA SE system.*

From Fig. 2 (a), there are decomposition, enhancement and reconstruction components involved in the DWPT-RPCA SE system. First, the DWPT performed in Eq. (1) is used to split a noise-corrupted signal $\mathbf{n}_0^0$ into $B$ sub-band signals $\mathbf{n}^J$ in the decomposition stage. Then, all of the sub-band signals are individually enhanced via RPCA-wise enhancement system. Finally, these updated sub-band signals are joined using the IDWPT defined in Eq. (2) to construct the enhanced signal $\mathbf{c}_0^0$.

For an arbitrary sub-band input signal, $\mathbf{n}_b^J$, in the Fig. 2 (b), the framing operation $h\{\cdot\}$ without any windowing (equivalent to using a rectangular window) is first performed to produce overlapping frames. These frames are then arranged in sequence as the columns for matrix $\mathbf{N}_b$. Thus, the $m$-th column

of $\mathbf{N}_b$ is

$$\mathbf{N}_b(:,m) = \begin{bmatrix} \mathbf{n}_b^J((m-1)P+1) \\ \mathbf{n}_b^J((m-1)P+2) \\ \vdots \\ \mathbf{n}_b^J((m-1)P+L) \end{bmatrix} \quad (8)$$

where $L$ and $P$ denote the frame size and shift, respectively. The gain matrix $\mathbf{G}_b$ in Eq. (6) is extracted from the input noisy matrix $\mathbf{N}_b$ by using the RPCA procedure described in Section 3. The OLA process as the de-framing scheme in Eq. (9) is applied to obtain the gain sequence $\mathbf{g}_b$ for the $b$-th sub-band signal $\mathbf{n}_b^J$. Then, the virtually clean signal component $\tilde{\mathbf{c}}_b^J$ is derived in Eq. (10) accordingly through the multiplication between the input sub-band signal $\mathbf{n}_b^J$ and the gain sequence $\mathbf{g}_b$. Notably, the gain vector $\mathbf{g}_b$ has the same size as the original sub-band signal $\mathbf{n}_b^J$.

$$c_b^J[n] = h^{-1}\{\mathbf{G}_b\} = \sum_{k=0}^{K-1} \mathbf{G}_b[k, n-kL], \quad (9)$$

where $n$ is the time index, $K$ is the total number of columns of $\mathbf{G}_b$, and $\mathbf{G}_b[i,j]$ is the *ij-th* element of $\mathbf{G}_b$.

$$\tilde{\mathbf{c}}_b^J = \mathbf{g}_b. \times \mathbf{n}_b^J. \quad (10)$$

Please note that the whole DWPT-RPCA SE system operates in the time domain of signals, including the stages of decomposition, enhancement, and reconstruction. Furthermore, the output $\mathbf{c}_0^0$ will be identical to the input $\mathbf{n}_0^0$ for the DWPT-RPCA SE in the case when no enhancement occurs, which indicates the direct concatenation of the front-end decomposition and the back-end reconstruction is distortion-less.

# 5. Experiments

## 5.1. Experimental setup

To evaluate the performance of DWPT-RPCA, we conduct the experiments on the MHINT sentences [22], which contains 300 utterances pronounced by a native Mandarin male speaker and recorded in a noise-free environment at a sampling rate of 16 kHz. These utterances were further down-sampled to 8 kHz for the experiments to form a more challenging task. From this database, 50 utterances were extracted as a clean set for evaluation. Eight types of noise (subway, exhibition, car, street, restaurant, babble, airport, and train-station) drawn from Aurora-2 [23] were artificially added to these clean utterances at six signal-to-noise ratios (SNRs) ranging from 20 to –5 dB with a 5-dB interval to generate the noisy testing set.

A 5-sample frame shift ($P = 5$) was used for DWPT-RPCA while the frame size and the level of the DWPT/IDWPT were varied, and the corresponding issues will be discussed in section 5.2. For comparison, the STFT-based RPCA and MMSE SE system was implemented in this study. Here, we applied the improved minima controlled recursive averaging [24] noise estimating technique to estimate noise conditions for MMSE. The SE scenarios were evaluated using: (1) the quality test for the hearing-aid speech quality index (HASQI) [25], (2) the perceptual test for the hearing-aid speech perception index (HASPI) [26], and (3) the objective test for segmental SNR improvement (SSNRI) and speech distortion index (SDI). Notably, the HASQI and HASPI have been proven to provide high correlation scores with human quality assessment and perception, and are developed to evaluate sound quality and perception for both hearing-impaired patients and people without hearing issues. The scores for the HASQI and HASPI both range from zero to one. Higher scores of the HASQI and HASPI correspond to better sound quality and intelligibility, respectively.

## 5.2. Performance evaluation

The average SSNRI scores over all noise conditions (eight noise types and six SNR levels) with respect to DWPT-RPCA SE with four different levels ($J = 0, 1, 2, 3$) of DWPT/IDWPT were shown in Table 1. The applied frame size for DWPT-RPCA in the experiments was 100 ($L = 100$). From the table, scores were increased along the level from one to three, and three-level provides the highest score of SSNRI. With applying three-level DWPT/IDWPT (decomposed a full-band signal into eight sub-bands sequences) to DWPT-RPCA, Table 2 listed average scores of SSNRI over all noise conditions in terms of varying frame size from 200- to 1000-samples with 200 samples an interval. From the table, we observed that SSNRI scores were increased along the frame length, and frame length with 1000-samples provides the best performance. The above observations show that high input feature dimensions with large number of sub-bands enable RPCA to more effectively suppress noise signals in the time domain. In the following experiments, 1000-samples frame length and three-level DWPT/IDWPT were used for DWPT-RPCA, and was denoted as "D-RPCA".

Next, we tested the SDI, HASQI and HASPI scores for all noise conditions for DWPT-RPCA, and compared the results with those from the noisy baseline, full-band MMSE and STFT-RPCA SE systems denoted as "Noisy", "MMSE" and "S-RPCA-F-256", respectively. The evaluation results were listed in Table 3. To achieve a fair comparison, 256- or 1000-samples frame length were used for STFT to create two different spectrograms for the input time signal. Each of the created spectro-

Table 1: *The average scores of SSNRI for DWPT-RPCA SE with four different levels of DWPT/IDWPT on the testing set.*

| $J$ | **0** | **1** | **2** | **3** |
|---|---|---|---|---|
| **SSNRI** | 0.2134 | 0.1947 | 0.3197 | ***0.6460*** |

Table 2: *The average scores of SSNRI for DWPT-RPCA with five different frame length on the testing set.*

| $L$ | **200** | **400** | **600** | **800** | **1000** |
|---|---|---|---|---|---|
| **SSNRI** | 1.0584 | 1.6392 | 2.2741 | 2.7066 | ***2.9949*** |

Table 3: *The average SDI, HASQI and HASPI scores on all testing set for Noisy, STFT-RPCA and DWPT-RPCA SE techniques.*

| **Index** | **SDI** | **HASQI** | **HASPI** |
|---|---|---|---|
| **Noisy** | 0.5543 | 0.1299 | 0.5899 |
| **MMSE** | 0.5434 | 0.1376 | 0.6495 |
| **S-RPCA-F-256** | 0.5198 | 0.1309 | 0.6169 |
| **S-RPCA-8-256** | 0.5349 | 0.1288 | 0.6030 |
| **S-RPCA-8-1000** | 1.9902 | 0.0147 | 0.0503 |
| **D-RPCA** | ***0.5047*** | ***0.1728*** | ***0.7716*** |

gram was divided along the frequency axis into eight equally-spaced sub-bands for STFT-RPCA SE. RPCA was then used to update each sub-band. These updated sub-band matrices were combined to form the enhanced spectrogram, and then transformed into the time domain with an inverse STFT (ISTFT) on the corresponding frame-length setting. In Table 3, these two sub-band-based STFT-RPCA SE were denoted as "S-RPCA-8-256" and "S-RPCA-8-1000", where "8" represents eight sub-band, and "256" and "1000" were the frame length. Notably, the applied frame size and shift for both "S-RPCA-F-256" and "S-RPCA-8-256" were 256 and 80 samples, respectively, while the 5-points frame shift was used for "S-RPCA-8-1000".

From Table 3, MMSE, S-RPCA-F-256, S-RPCA-8-256, and D-RPCA can effectively decrease the signal distortion and improve sound quality and intelligibility simultaneously from noisy inputs. In addition, S-RPCA-F-256 shows better results on all the evaluation matrices than those from S-RPCA-F-256 and S-RPCA-8-1000, indicating that the performance of RPCA-based SE systems can further be improved by carefully selecting the input dimension. Furthermore, the proposed DWPT-RPCA SE provides the best system performance on the evaluations of SDI, HASQI and HASPI scores, implying that the DWPT/IDWPT can provide better signal bases for RPCA and lead to less distortion of transformation relative to STFT/ISTFT. Here, the notation "S-RPCA-F-256" was simplified to "S-RPCA" in the following comparisons.

For Noisy, MMSE, S-RPCA and D-RPCA, we evaluated the HASQI and HASPI scores, which were averaged over six SNR conditions of each noise environment, in Tables 4 and 5, respectively. From both tables, we observed that S-RPCA, MMSE and D-RPCA consistently improve the sound quality and intelligibility from noisy signals in all noise environments. In addition, D-RPCA showed the best score in each noise type

Table 4: *Evaluated results on HASQI in eight noise environments.*

| Method | Noisy | MMSE | S-RPCA | D-RPCA |
|---|---|---|---|---|
| **Subway** | 0.1183 | 0.1286 | 0.1242 | ***0.1502*** |
| **Exhibition** | 0.1336 | 0.1409 | 0.1375 | ***0.1614*** |
| **Car** | 0.1158 | 0.1369 | 0.1218 | ***0.1712*** |
| **Street** | 0.1216 | 0.1256 | 0.1235 | ***0.1657*** |
| **Restaurant** | 0.1330 | 0.1334 | 0.1314 | ***0.1714*** |
| **Babble** | 0.1309 | 0.1346 | 0.1263 | ***0.1768*** |
| **Airport** | 0.1446 | 0.1452 | 0.1405 | ***0.1917*** |
| **Train Station** | 0.1419 | 0.1557 | 0.1417 | ***0.1937*** |

Table 5: *Evaluated results on HASPI in eight noise environments.*

| Method | Noisy | MMSE | S-RPCA | D-RPCA |
|---|---|---|---|---|
| **Subway** | 0.5715 | 0.6218 | 0.6133 | ***0.7063*** |
| **Exhibition** | 0.6702 | 0.7168 | 0.7028 | ***0.7744*** |
| **Car** | 0.5174 | 0.6249 | 0.5575 | ***0.7569*** |
| **Street** | 0.5525 | 0.6026 | 0.5789 | ***0.7455*** |
| **Restaurant** | 0.6049 | 0.6399 | 0.6228 | ***0.7685*** |
| **Babble** | 0.5749 | 0.6239 | 0.5827 | ***0.7754*** |
| **Airport** | 0.6170 | 0.6713 | 0.6384 | ***0.8214*** |
| **Train Station** | 0.6111 | 0.6945 | 0.6385 | ***0.8239*** |

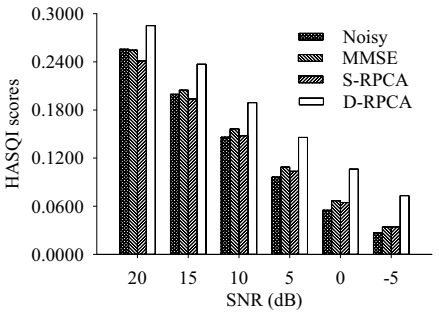



Figure 3: *Evaluated results on HASQI at six SNR conditions.*

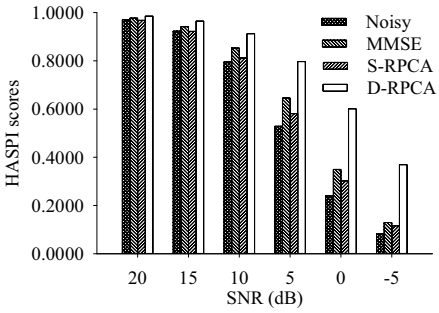



Figure 4: *Evaluated results on HASPI at six SNR conditions.*

when comparing with those from S-RPCA and MMSE. The highest score demonstrated that the speech quality and intelligibility can further be improved by enhancing magnitude and phase components simultaneously in a time signal.

Figs. 3 and 4 compared Noisy, MMSE, S-RPCA and D-RPCA in terms of the averaged HASQI and HASPI metric scores over all noise types at six SNRs, respectively. Both figures showed that D-RPCA outperforms the MMSE and S-RPCA SE system when considering all the HASQI and HASPI scores on different SNR conditions. This observation confirms again that DWPT-RPCA SE provides a better sound quality as well as a higher intelligibility than STFT-RPCA SE.

## 6. Conclusion

We proposed a novel subspace-based SE technique integrating DWPT/IDWPT and RPCA. Instead of constructing the subspaces on the magnitude spectrogram, time-domain subspaces were created for RPCA to enhance noisy input. In addition, the well-defined filters in DWPT/IDWPT allow distortion-less decomposition/reconstruction for the input signal, and the underlying sub-band signals can be enhanced using RPCA. Results showed that DWPT-RPCA results in less distortion of enhanced speech utterances than STFT-based RPCA and MMSE, thereby improving sound quality and intelligibility. Experiments incorporating DWPT/IDWPT with model-based SE techniques such as deep neural network-wise and MMSE-STSA will be examined to determine if further improvement can be achieved.

## 7. Acknowledgements

The authors appreciate the financial support provided by Ministry of Justice (Grant number:106-1301-05-04-01).

# 8. References


[1] W. Hartmann, A. Narayanan, E. Fosler-Lussier, and D. Wang, "A direct masking approach to robust ASR," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 1993–2005, 2013.

[2] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed mwf-based noise reduction algorithms for binaural hearing aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 38–51, 2009.

[3] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013.

[4] P. Scalart *et al.*, "Speech enhancement based on a priori signal to noise estimation," in *Proc. ICASSP*, pp. 629–632, 1996.

[5] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," *Speech communication*, vol. 50, no. 6, pp. 453–466, 2008.

[6] P. C. Loizou, "Speech enhancement based on perceptually motivated bayesian estimators of the magnitude spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 857–869, 2005.

[7] Y. Tsao and Y.-H. Lai, "Generalized maximum a posteriori spectral amplitude estimation for speech enhancement," *Speech Communication*, vol. 76, pp. 112–126, 2016.

[8] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: A regenerative approach," *Speech Communication*, vol. 10, no. 1, pp. 45–57, 1991.

[9] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 334–341, 2003.

[10] Y. Benabderrahmane, S. A. Selouani, and D. O'Shaughnessy, "Blind speech separation for convolutive mixtures using an oriented principal components analysis method," in *Proc. EUSIPCO*, pp. 1553–1557, 2010.

[11] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using nonlinear PCA type learning," *Neural networks*, vol. 7, no. 1, pp. 113–127, 1994.

[12] K. Hermus, P. Wambacq, *et al.*, "A review of signal subspace speech enhancement and its application to noise robust speech recognition," *EURASIP Journal on applied signal processing*, vol. 2007, no. 1, pp. 195–195, 2007.

[13] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in *Proc. ICASSP*, pp. 57–60, 2012.

[14] S.-S. Wang, A. Chern, Y. Tsao, J.-W. Hung, X. Lu, Y.-H. Lai, and B. Su, "Wavelet speech enhancement based on nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 23, no. 8, pp. 1101–1105, 2016.

[15] N. A. Whitmal, J. C. Rutledge, and J. Cohen, "Reducing correlated noise in digital hearing aids," *IEEE engineering in medicine and biology magazine*, vol. 15, no. 5, pp. 88–96, 1996.

[16] M. V. D. Shinde, M. C. Patil, and M. S. D. Ruikar, "Wavelet based multi-scale principal component analysis for speech enhancement," *matrix*, vol. 4, p. 2, 2012.

[17] A. Bouzid, N. Ellouze, *et al.*, "Speech enhancement based on wavelet packet of an improved principal component analysis," *Computer Speech & Language*, vol. 35, pp. 58–72, 2016.

[18] C. Sun, Q. Zhang, J. Wang, and J. Xie, "Noise reduction based on robust principal component analysis," *Journal of Computational Information Systems*, vol. 10, no. 10, pp. 4403–4410, 2014.

[19] M. Gokhale, D. K. Khanduja, *et al.*, "Time domain signal analysis using wavelet packet decomposition approach," *Int'l J. of Communications, Network and System Sciences*, vol. 3, no. 03, p. 321, 2010.

[20] D. D. Ariananda, M. K. Lakshmanan, and H. Nikookar, "An investigation of wavelet packet transform for spectrum estimation," *arXiv preprint arXiv:1304.3795*, 2013.

[21] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. NIPS*, pp. 612–620, 2011.

[22] L. L. Wong, S. D. Soli, S. Liu, N. Han, and M.-W. Huang, "Development of the mandarin hearing in noise test (MHINT)," *Ear and hearing*, vol. 28, no. 2, pp. 70S–74S, 2007.

[23] H.-G. Hirsch and D. Pearce, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*, 2000.

[24] I. Cohen, "Speech enhancement using a noncausal a priori snr estimator," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 725–728, 2004.

[25] J. M. Kates and K. H. Arehart, "The hearing-aid speech quality index (HASQI)," *Journal of the Audio Engineering Society*, vol. 58, no. 5, pp. 363–381, 2010.

[26] J. M. Kates and K. H. Arehart, "The hearing-aid speech perception index (HASPI)," *Speech Communication*, vol. 65, pp. 75–93, 2014.